OHDSI内では、実名での活動になります。
Zoom参加時も「名前は実氏名で」お願いします。

OHDSI Japan evening conference #8

イブニング カンファレンス(第8回)
2020.7.28

2020 OHDSI Symposium

Global&USのシンポジウム（メインのもの）

- 10月18日～21日(４日間）、WEB開催。
- 参加費無料～（寄付も受け付けます）
- 参加登録受付中↓

https://www.ohdsi.org/2020-symposium-registration/

Asia-Pacific Symposium 2020

- Asia-Pacificのシンポジウム（今回初めて）
- 12月5日～6日(半日2日間）、WEB開催。
- 参加費無料～のはず。

OHDSI Japanの活動場所

ここしばらくは
ネット上での活動。
そのうちリアルも
復活させたい。

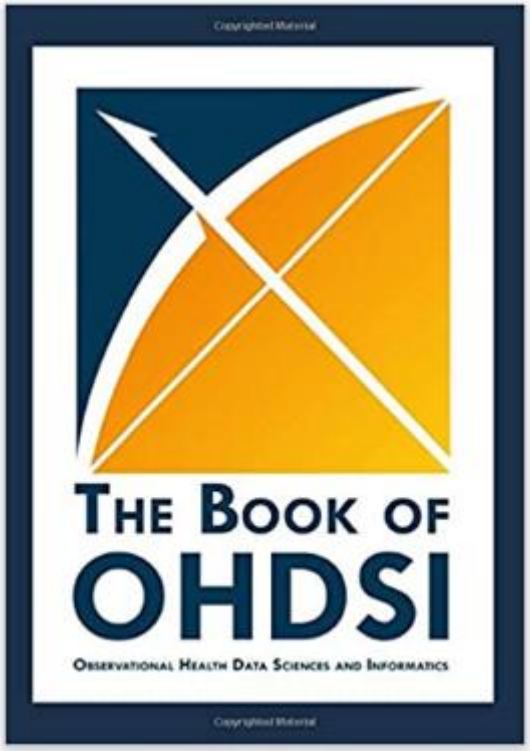
- OHDSI Japanサイト: 情報提供
  www.ohdsi-japan.org
- 本家Forumの日本コーナー
  Forumに登録すると、書き込める。世界中だれでも読める。
  https://forums.ohdsi.org/c/ohdsi-in-japan
- 日本コミュニティ メンバー登録
  Doorkeeperのサイト。ODJで一番大きな枠組み。案内メールが流れてくる。発言場所はない。イベントの参加登録ができる。
- 日本のコラボレータML
  メーリングリスト。ML内で発言できる。流量は多くない。登録希望者は平松まで（サイト/DKのお問い合わせから）
- 日本のSlack
  登録希望は岡田先生まで（お問い合わせでも転送します）。
  OHDSI Book翻訳のために作られたが、それ以外でもgenealチャンネルにどうぞ。WS内で発言できる。

The Book of OHDSI紹介

- OHDSIのすべて（？）が書かれた本
- 無料で読める
  https://ohdsi.github.io/TheBookOfOhdsi/

- Amazonで物理本も買える
  ($10.82+送料 / ¥2,007)

- 日本語版はOHDSI Japan
  メンバーで翻訳進行中



The Book of OHDSI紹介

第 1 部 The OHDSI Community

1. The OHDSI Community
2. Where To Begin
3. Open Science

第 2 部 Uniform Data Representation

4. The Common Data model
5. Standardized Vocabularies
6. Extract Transform Load

第 3 部 Data Analytics

7. Data Analytics Use Cases
8. OHDSI Analytics Tools
9. SQL and R
10. Defining Cohorts
11. Characterization
12. Population-Level Estimation
13. Patient-Level Prediction

第 4 部 Evidence Quality

14. Evidence Quality
15. Data Quality
16. Clinical Validity
17. Software Validiy
18. Method Validity

第 5 部 OHDSI Studies

19. Study Steps
20. OHDSI Network Research

Appendix

A. Glossary
B. Cohort definitions
C. Negative controls
D. Protocol template
E. Suggested Answers

References

7. Data Analytics Use Cases (1)

● ３つの主要なカテゴリがあります。

特徴：個人の特性またはデータベース全体、医療の実践に関する質問へ回答しようとするもの。
（例）新たに心房細動と診断された患者では、何人がワルファリンの処方を受けますか？

人口レベルの推定：集団における因果関係（アウトカムの要因）を推定するもの。
（例）心房細動と新たに診断された患者の場合、治療開始後の最初の年に、ワルファリンはダビガトランよりも大きな出血を引き起こしますか？

患者レベルの予測：特定の患者の予後を予測するもの
（例）大うつ病性障害と新たに診断された特定の患者について、患者が診断後の最初の1年間に自殺を試みる確率はどのくらいですか？

予測するものであって、因果関係をsuggestするものではない。

7. Data Analytics Use Cases (2)

●高血圧をテーマとしたときの例

●診療情報を利用する研究の限界

•プラセボと比較した介入の因果関係。

•市販薬に関連するもの。

•多くの結果やその他の変数があまり記録されていない。死亡率、行動転帰、ライフスタイル、および社会経済的地位など。

•患者は体調が悪いときにのみ受診するため、治療の効果を測定することは難しい場合がある。

•誤ったデータ、本来あるのに欠損しているデータもある。

8. OHDSI分析ツール

無料のODHSI分析ツールが利用できる。
もちろん使わないこともできる。

- ATLAS
  統合分析ツール
- メソッドライブラリ
  Rのライブラリ
- セットアップ方法
  Broadsea：Dockerイメージでツールを提供
  AWS：OHDSI-in-a-Box(小規模) と OHDSIonAWS(大規模)
  日本では「医療情報システムの安全管理に関するガイドライン」

9. SQLとR

●SqlRenderパッケージ

SQLにも方言があり、1つの標準の方言から、サポートされている方言のいずれかに変換できるRパッケージ。標準の方言（OHDSI SQL）は、主にMS SQL ServerにおけるSQL方言のサブセット

●DatabaseConnectorパッケージ

JavaのJDBCドライバーを使用して多くのデータベースプラットフォームに接続できる1つのRパッケージ。

●CDMに対するクエリ

●クエリ時のボキャブラリの使用

●QueryLibrary： CDMに対するクエリのライブラリ

●簡単な調査の設計

●SQLとRを使用した調査の実装

10. コホートの定義

- ATLASを使うときに多用する
- 診療情報RWDは、研究目的で収集されたものではないため、関心のある臨床データ要素を明示的に取得できない場合がある。関心事項を取得するロジックが必要 (Phenotyping)。
- ODHSIでは「1つ以上の包含基準を一定期間満たしている患者群」をコホートと呼ぶ。
- ３つの基準：コホートエントリイベント、包含基準、コホート離脱基準
- ルールベースコホート、確率的コホート
- コホート定義の妥当性
- 高血圧を例にしたコホートの定義
- ATLASを使用したコホートの実装
- SQLを使用したコホートの実装

TECHNICAL ADVANCE Open Access

The Generalized Data Model for clinical research

Mark D. Danese*, Marc Halperin, Jennifer Duryea and Ryan Duryea

- OMOP-CDMはETL大変なので、もっと簡単にしたいよね。
- うちはGDMってのにしているよ。

https://github.com/outcomesinsights/generalized_data_model

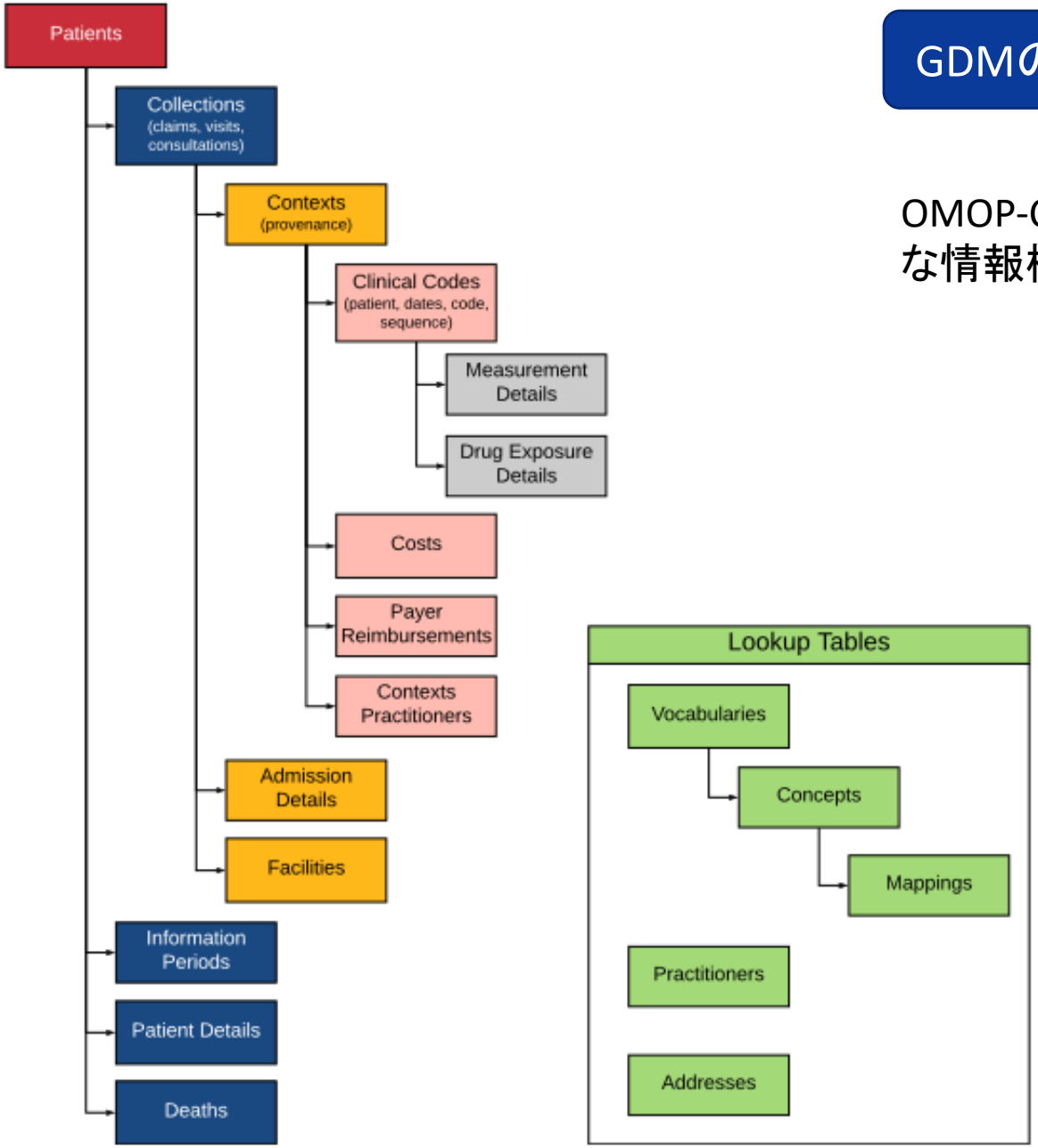
特に
- Native codeを使いたいとき
- Network studyをするつもりはないとき。

にはいいんじゃない？？

という話題。

GDMのテーブル構成

OMOP-CDMになりかけ？
な情報構成



GDM つづき

でも、Native codeでnetwork studyも実施しないのなら、そもそも何かのデータモデルに変換する必要なくない？

⇒ いや、組織内ではモデルを統一して同じクエリーを使えるようにしたかったんだよ。

ここで開発者登場！
- OMOP-CDMへのETLの簡易中間段階を作ることで楽に行いたくて始めた。
- 実は OMOP-CDMだけでなくSentinel-CDMへの変換も必要だった。

「コードMappingをしない中間段階モデル」
を経由して**複数モデル**へETLする。

OMOP化する利点４つ

(1)国際的なコミュニティにより専門家のアドバイスが得られ、成果が利用できる。（ほぼすべてのCDMはdomesticで、真に国際的なのはOMOP/OHDSIのみ）

(2)複数のソースデータに対して、同じqueryが使える。

(3)無料のODHSIツールが使える。

(4)OHDSI network study に参加できる。

(2)はGDMでもできる。Mappingが必要ならAthenaから持ってきてすれば良いので。

(1)もある程度は可能？ GDMはOMOPからかなりのアイディアを持ってきているため。

OMOPなくてはGDM存在できない…

データモデルの関連資料

A common data model in Europe?
– Why? Which? How?

- 2017年12月にロンドンで開かれた会合
- 「医薬品規制の判断に役立つ欧州でのデータモデル」について議論
- 資料は下記URLから

https://www.ema.europa.eu/en/events/common-data-model-europe-why-which-how

コラボレータMLの話題から

ICD9CM、ICD10CM、ICD9、ICD10これらのコード間の関係性を知りたい場合、皆さんはどのような方法で検索していますでしょうか。
効率的な方法を探し中です。良い方法ありましたら共有頂けるとありがたいです。

ちなみに、Athenaでは疾患/症状徴候に関するコードはSNOMEDを中心とした各コード間の関連が分かるというものであって、例えば上記のようなICDコード間の関係性を見るには適していないとの理解ですがあっていますでしょうか？
https://athena.ohdsi.org/search-terms/start

コラボレータMLの話題から

ICDコード間の変換は1:1で実施できるものではないですが、それぞれのコードの定義が扱いやすく電子化されているものとしては、Rの icd パッケージがあります。詳細は https://jackwasey.github.io/icd/ にありますが、こちらを使うと、それぞれのコードの説明をパターンマッチすることは以下のように可能です。パッケージに含まれているicd10cm2019データの25行目にマッチするicd9のコードを調べてみたら、icd9cm2014_leafデータの14行目にあったことがわかります。

```
> library(icd)
> grep(paste0(".*", icd10cm2019[25,"short_desc"], ".*"), icd9cm2014_leaf$short_desc)
[1] 14

> icd9cm2014_leaf[14,]
    code           short_desc            long_desc
14 00323 Salmonella arthritis Salmonella arthritis

> icd10cm2019[25,]
    code billable           short_desc            long_desc
25 A0223     TRUE Salmonella arthritis Salmonella arthritis
three_digit                                   major
A02                   Other salmonella infections
                       sub_chapter                                          chapter
25 Intestinal Infectious Diseases     Certain infectious and parasitic diseases
```